

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-155012

(43)Date of publication of application : 08.06.2001

(51)Int.Cl. G06F 17/21  
G06F 12/00  
G06F 17/30

(21)Application number : 2000-297861

(71)Applicant : RICOH CO LTD

(22)Date of filing : 29.09.2000

(72)Inventor : KURT PIASORU

(30)Priority

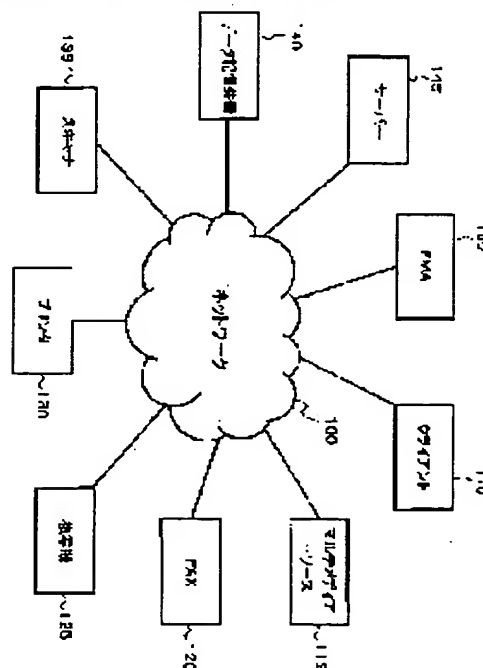
Priority number : 1999 410364 Priority date : 30.09.1999 Priority country : US

**(54) METHOD, DEVICE AND SYSTEM FOR MANAGING DOCUMENT IN NETWORK**

(57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a simple and clear interface for reducing the intervention of a user for managing documents generated from various sources, and efficiently retrieving an electronic file.

**SOLUTION:** A file managing device (FMA) 150 carries out a fetching function, an indexing function and a retrieving function for electronic documents in the environment connected to a network 100 together with a client 110, facsimile equipment 120, copy machine 125, printer 130, scanner 135, data storage device 140 and server 145 or the like and in order to facilitate the interaction of the user and the device, the graphical user interface of web base is provided.

**LEGAL STATUS**

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision ...]

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開2001-155012

(P2001-155012A)

(43) 公開日 平成13年 6 月 8 日 (2001. 6. 8)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/21	5 7 0	G 0 6 F 17/21	5 7 0 D
12/00	5 2 0	12/00	5 2 0 J
17/30	1 1 0	17/30	1 1 0 F
	1 7 0		1 7 0 A
	2 3 0		2 3 0 Z

審査請求 未請求 請求項の数29 O L (全 20 頁)

(21) 出願番号 特願2000-297861 (P2000-297861)

(22) 出願日 平成12年 9 月29日 (2000. 9. 29)

(31) 優先権主張番号 0 9 / 4 1 0 3 6 4

(32) 優先日 平成11年 9 月30日 (1999. 9. 30)

(33) 優先権主張国 米国 (U S)

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込 1 丁目 3 番 6 号

(72) 発明者 カート ピアソル

アメリカ合衆国 カリフォルニア州

94025 メンローパーク スイート 115

サンド ヒル ロード 2882 リコーコー

ポレーション内

(74) 代理人 100073760

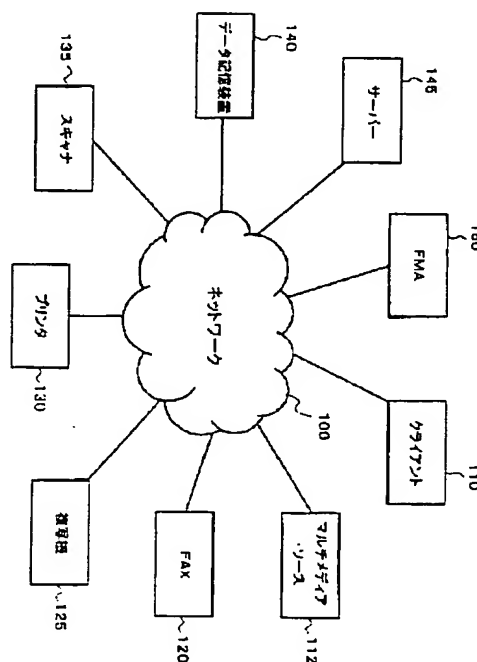
弁理士 鈴木 誠 (外 1 名)

(54) 【発明の名称】 ネットワーク内の文書を管理するための方法、装置及びシステム

(57) 【要約】

【課題】 様々なソースによって生成された文書を管理するためのユーザの介入を減らし、また、効率的な電子ファイル検索のための単純明解なインターフェースを提供する。

【解決手段】 ファイル管理装置 (FMA) 150 は、クライアント 110、ファクシミリ装置 120、複写機 125、プリンタ 130、スキャナ 135、データ記憶装置 140、サーバー 145 などとともにネットワーク 100 に接続された環境で、電子文書の取り込み機能、索引付け機能、検索機能を遂行し、また、ユーザと装置の対話を容易にするため、ウェブ・ベースのグラフィカル・ユーザ・インターフェースを提供する。



## 【特許請求の範囲】

【請求項1】 ユーザ・インターフェース、ネットワークに接続された第1の装置より取り込まれた文書を受信するために利用可能なネットワーク・インターフェース、前記取り込まれた文書を順序づけるための1つ以上の待ち行列、前記1つ以上の待ち行列を管理するための1つ以上のデーモン、及び前記1つ以上の待ち行列を格納するための記憶媒体、を有することを特徴とするネットワーク内の文書を管理するための装置。

【請求項2】 前記ユーザ・インターフェースがグラフィカル・ユーザ・インターフェースであることを特徴とする請求項1記載の装置。

【請求項3】 前記グラフィカル・ユーザ・インターフェースが、前記ネットワーク・インターフェースを通じてアクセス可能なCGIスクリプトを含むことを特徴とする請求項2記載の装置。

【請求項4】 前記グラフィカル・ユーザ・インターフェースがウェブ・ベースの表示ページを生成するために利用できることを特徴とする請求項2記載の装置。

【請求項5】 前記ネットワーク・インターフェースにより1つ以上の取り込み装置にネットワーク経由で通信可能に接続されることを特徴とする請求項1記載の装置。

【請求項6】 前記ネットワーク・インターフェースがFTPによるデータの送受信に利用できることを特徴とする請求項5記載の装置。

【請求項7】 前記1つ以上の取り込み装置がネットワーク化オフィス装置であることを特徴とする請求項5記載の装置。

【請求項8】 前記1つ以上の待ち行列中の第1の待ち行列が少なくとも1つの文書ディレクトリを含むことを特徴とする請求項1記載の装置。

【請求項9】 前記少なくとも1つの文書ディレクトリが1つ以上のコンテンツ・ファイルを含むことを特徴とする請求項8記載の装置。

【請求項10】 前記少なくとも1つの文書ディレクトリが前記取り込まれた文書の少なくとも1部を格納したメタデータ・ファイルを含むことを特徴する請求項8記載の装置。

【請求項11】 前記記憶媒体が当該装置上の全ての文書の位置を記録するマスター文書リストを含むことを特徴とする請求項1記載の装置。

【請求項12】 ネットワークと通信可能に接続され、データを取り込むように構成された第1の装置、及び前記ネットワークと通信可能に接続された第2の装置を有し、前記第2の装置は、ユーザ・インターフェース、

前記第1の装置より少なくとも一部の取り込みデータを受信するように構成されたネットワーク・インターフェース、

前記少なくとも一部の取り込みデータを順受けするための1つ以上の待ち行列、

前記1つ以上の待ち行列を管理するための1つ以上のデーモン、及び前記1つ以上の待ち行列を記憶するための記憶媒体を有する。ことを特徴とするネットワーク内の文書を管理するためのシステム。

【請求項13】 前記ネットワーク・インターフェースがFTPにより取り込みデータを受信することを特徴とする請求項12記載のシステム。

【請求項14】 前記第1及び第2の装置の少なくとも一方がネットワーク化オフィス装置であることを特徴とする請求項12記載のシステム。

【請求項15】 前記第1及び第2の装置の両方がネットワーク化オフィス装置であることを特徴とする請求項12記載のシステム。

【請求項16】 前記少なくとも一部の取り込みデータが文書の書誌的情報を含むことを特徴とする請求項12記載のシステム。

【請求項17】 前記1つ以上の待ち行列が1つ以上の文書ディレクトリのリストを含むことを特徴とする請求項12記載のシステム。

【請求項18】 前記ユーザ・インターフェースが、リスト表示、カレンダー表示、サムネイル表示、ストリーム表示を含む複数のグラフィカル文書画面表示を有するグラフィカル・ユーザ・インターフェースからなることを特徴とする請求項12記載のシステム。

【請求項19】 ネットワーク内の文書を管理するための方法であって、第1の装置でデータを取り込み；かつ前記取り込まれたデータの少なくとも一部を第2の装置へ透過的に伝達し、

前記第2の装置が、前記取り込まれたデータの前記少なくとも一部を1つ以上の待ち行列に順序付けし、

前記1つ以上の待ち行列に対し1つ以上の処理操作を実行し、

前記取り込まれたデータに関する1つ以上の問い合わせを受け付け、かつ前記1つ以上の問い合わせに対し問い合わせ結果を表示することにより応答する、ことを特徴とする方法。

【請求項20】 前記データがFTPにより取り込まれることを特徴とする請求項19記載の方法。

【請求項21】 前記第2の装置が、ディレクトリを生成して該ディレクトリに前記取り込まれたデータの前記少なくとも一部を複写することにより、前記取り込まれたデータの前記少なくとも一部を順序付けすることを特徴とする請求項19記載の方法。

【請求項22】 前記第2の装置が、前記取り込まれたデータの前記少なくとも一部の取り込みデータを前記ディレクトリの名前として使用することを特徴とする請求項21記載の方法。

【請求項23】 前記処理操作に光学文字認識、索引付け、圧縮及びバックアップが含まれることを特徴とする請求項19記載の方法。

【請求項24】 前記1つ以上の問い合わせがCGIスクリプトを使用して実現されることを特徴とする請求項19記載の方法。

【請求項25】 前記問い合わせ結果が1つ以上の表示形式でグラフィック表示されることを特徴とする請求項19記載の方法。

【請求項26】 前記第2の装置が前記取り込まれたデータの前記少なくとも一部を索引付けすることを特徴とする請求項19記載の方法。

【請求項27】 前記取り込まれたデータの全てを第3のデバイスに伝達して格納することを特徴とする請求項19記載の方法。

【請求項28】 前記問い合わせ結果が複数の表示形式の中から選択可能な1つの表示形式でグラフィック表示されることを特徴とする請求項19記載の方法。

【請求項29】 前記複数の表示形式にリスト表示、カレンダー表示、サムネイル表示、ストリーム表示が含まれることを特徴とする請求項28記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、電子文書の管理に係り、特に、ネットワーク内の電子文書の管理のための方法、装置及びシステムに関する。

【0002】

【従来の技術】ローカルエリアネットワーク(LAN)、広域ネットワーク(WAN)から家庭内ネットワークにおよぶデジタル通信ネットワークの最近の普及により、電子装置の通信可能な経路が数多く生まれた。ネットワーク化が可能な装置の仲間は、もはやコンピュータだけではなく、例えばプリンタ、複写機、さらにはファクシミリ装置などの装置を含むまでに拡大した。

【0003】ネットワーク化装置の数は増加の一途を辿り、それら装置により処理されるデータ量も増加が続いている。オフィス環境にいる人が毎日、大量のデータを電子的データ及び印刷物の両方の形式で送受信することは、そう珍しいことではない。

【0004】人によって送受信されるデータの大部分が、電子メール(eメール)になりつつある。普通、人は電子メールを受信すると、それを読んでから、メッセージのコピーを保存するか削除するか決める。電子メールを特定部分だけを読んで削除してしまい、後日、その電子メールを保存しておくべきであったことに気がつくことがよくある。同様に、電子メールのメッセージを保存

する人が、ある期間、1つ以上のメッセージを間違った場所に保存することもある。その人が配属されているオフィスにシステム管理者がいるならば、そのメッセージを再び見つけ出してもらうことができるかもしれない。しかし、困ったことに、特定のメッセージを突き止めようとしても、それが削除されてしまってからでは困難であり、また、間違った場所に保存したメッセージを探し出すには時間がかかってしまうこともある。

【0005】ワープロ・ファイル、スプレッド・シート、画像ファイルなどの電子データは、削除されても間違った場所に保存されても問題になりにくい。この種のファイルは、生成した後に他の人に電子データとして送信されることがあり、また、プリントアウトされるのがごく普通であろう。あるファイルのドラフト・バージョンの作成に携わる人は、色々な時点で、そのファイルの様々なバージョンをプリントアウトするのが普通である。人は、ファイルのある旧バージョンをうっかり削除してしまってから、削除したバージョンより保存されているバージョンのほうが古いバージョンであることに気付くことがある。

【0006】

【発明が解決しようとする課題】よって、ユーザの介入を減らし、できればユーザの介入を最小限度にして、様々なソースによって生成された電子ファイル及び文書进行管理し、効率的な電子ファイル検索のための単純明解なインターフェースを提供するシステムの実現が望まれる。本発明の目的は、そのような要求に応えるための手段を提供することにある。

【0007】

【課題を解決するための手段】上記目的を達成するための本発明の装置の主要な特徴は、請求項1記載のように、ユーザ・インターフェース、ネットワークに接続された第1の装置より取り込まれた文書を受信するために利用可能なネットワーク・インターフェース、前記取り込まれた文書を順序づけるための1つ以上の待ち行列、前記1つ以上の待ち行列を管理するための1つ以上のデーモン、及び、前記1つ以上の待ち行列を格納するための記憶媒体を有することである。

【0008】また、前記目的を達成するための本発明のシステムの主要な特徴は、請求項12記載のように、ネットワークと通信可能に接続され、データを取り込むように構成された第1の装置と、前記ネットワークと通信可能に接続された第2の装置とを有し、前記第2の装置が、ユーザ・インターフェースと、前記第1の装置より少なくとも一部の取り込みデータを受信するように構成されたネットワーク・インターフェースと、前記少なくとも一部の取り込みデータを順受けするための1つ以上の待ち行列と、前記1つ以上の待ち行列を管理するための1つ以上のデーモンと、前記1つ以上の待ち行列を記憶するための記憶媒体とを有することである。

【0009】また、前記目的を達成するための本発明の方法の主要な特徴は、請求項19記載のように、第1の装置でデータを取り込み、前記取り込まれたデータの少なくとも一部を第2の装置へ透過的に伝達するとともに、前記第2の装置が、前記取り込まれたデータの前記少なくとも一部を1つ以上の待ち行列に順序付けし、前記1つ以上の待ち行列に対し1つ以上の処理操作を実行し、前記取り込まれたデータに関する1つ以上の問い合わせを受け付け、前記1つ以上の問い合わせに対し問い合わせ結果を表示することにより応答することである。

【0010】以上に述べた本発明の特徴及びその他の本発明の特徴は、以下の実施の形態に関連した説明により明白となる。

【0011】

【発明の実施の形態】以下、ネットワーク内の文書（電子文書）の管理のための方法、装置及びシステムについて説明する。以下の記述中に、あくまで説明のためのものであるが、本発明を完全に理解できるように様々な具体例が示される。しかし、当業者には、そのような具体例によることなく本発明を実施し得ることは明白であろう。他方、本発明を分かりにくくしないため、周知の構造や装置はブロック図の形で表現される。

【0012】以下の詳細な説明には、コンピュータ・メモリ内のデータビットに対する操作のアルゴリズム及び記号表現によって表された部分がある。このようなアルゴリズム記述及び表現は、データ処理技術分野において、当業者が研究内容を他の当業者に最も効率的に伝えるために用いる手段である。あるアルゴリズムがあり、それが概して期待した結果に至る筋の通ったステップの系列だと理解されたとする。これらステップは、物理量の物理的処理を要するステップである。必ずという訳ではないが、これらの物理量は記憶、転送、結合、比較、その他処理が可能な電気的または磁気的信号の形をとるのが普通である。これらの信号をビット、値、要素、記号、文字、用語、数値等で表わすのが、主に慣用上の理由から便利な場合があることが分かっている。

【0013】しかしながら、このような用語や同様の用語は、適切な物理量に関連付けられるべきであり、また、これら物理量に付けた便宜上のラベルに過ぎないということに留意すべきである。以下の説明から明らかのように、特に断わらない限り、“処理”、“演算”、“計算”、“判定”、“表示”等の用語によって論じられることは、コンピュータシステムのレジスタ及びメモリの内部の物理的（電子的）な量として表現されたデータを処理して、コンピュータシステムのメモリやレジスタ、その他同様の情報記憶装置、情報伝送装置又は表示装置の内部の同様に物理量として表現された他のデータへ変換する、コンピュータシステムや同様の電子演算装置の作用及びプロセスを意味する。

【0014】本発明は、ここに述べる操作を実行するた

めの装置にも関係するものである。このような装置は、所要目的のために専用に作られてもよいし、内蔵のコンピュータ・プログラムにより選択駆動もしくは再構成された汎用コンピュータであってもよい。そのようなコンピュータ・プログラムは、コンピュータが読み取り可能な記憶媒体、限定するわけではないが例えば、フロッピー（登録商標）ディスク、光ディスク、CD-ROM、光磁気ディスクなどの任意の種類のディスク、リードオンリーメモリ（ROM）やランダムアクセスメモリ（RAM）、EPROM、EEPROM、磁気カード又は光カードなど、すなわち電子的命令の記憶に適したコンピュータのシステムバスに接続された任意種類の媒体に格納してもよい。ここに提示するアルゴリズム及び表示は、本質的に、いかなる特定のコンピュータ、その他の装置とも関わりがない。様々な汎用マシンを、ここに述べる内容に従ったプログラムで使用し得るが、所要の方法のステップの実行のために、より特化した装置を作るほうが好都合であるかもしれない。これら多様なシステムに要求される構造は以下の説明から明らかになる。また、特定のプログラム言語と関連させて本発明を説明することがあるが、ここに述べる本発明の内容を実現するために様々なプログラミング言語を使用し得ることが理解されよう。

【0015】処理操作の全部又は一部を、コンピュータ又は専用装置の1つ以上の処理装置（例えばCPU）上で動作するソフトウェアにより実行してもよいが、デジタル・ロジック及び／又は回路、集積回路（例えばASIC）あるいは他の半導体基板で実行してもよい。

【0016】《システムの概要》ファイル管理装置（“FMA” file management appliance）は、複数のプロセス及び待ち行列を利用して文書取り込みサービス及び索引付けサービスをネットワークに提供する装置である。一実施例では、FMAは、ペアド・ダウン（pared-down）UNIX（登録商標）ベース・オペレーティングシステムを使用したシン・サーバーとして実現される。別の実施例では、FMAは、様々なソースを通じて取り込まれた文書の集合のためのウェブ・ベース・インターフェースをユーザに提供する。さらに別の実施例では、FMAは、ウェブ・ベースCGI（common gateway interface）スクリプトを使って標準化したユーザ入力及び表示を可能にする。エンド・ユーザにとって、FMAはローカル・ネットワーク用の検索エンジンのように見える。共有ネットワーク上の他の装置にとっては、FMAは文書取り込みを可能にする装置である。この文書取り込みには、標準的なインターネット・プロトコルを利用できる。FMAは、ネットワーク化オフィス装置（NOA: networked office appliance）であるか、あるいは、NOAアーキテクチャの一部である。NOAと、それをサポートするためのアーキテクチャの両方について、“Method and Apparatus for Automatic Network Co

“nfigurations”の表題で1998年11月12日に出願され、本出願の出願人に譲渡された、同時係属出願第09/191,277号に説明されている。同出願をここに援用する。

【0017】簡単に説明すると、一実施例では、NOAアーキテクチャは、オフィスやネットワーク管理者のいない環境で、環境設定もしくはセットアップを必要とせずに、ネットワークへの機器の増設を可能にする。同時に、NOAアーキテクチャは、標準的に管理されたネットワークに対し、既存のネットワーク装置との衝突を起こさせることなく機器の容易な増設を可能にする。さらに、複数のNOAは、ネットワーク化環境で密に協調し、ネットワークに接続された時に自動的に情報を検出して互いに共有することができる。

【0018】NOAアーキテクチャは、例えばTCP/IPネットワークなどのネットワーク上の装置を自動的に環境設定するための仕組みを提供する。提供し得る具体的な環境設定方法には、ネットワーク・アドレス割り当て方法、ドメイン・ネーム・サービス(DNS)データベース登録方法、ネットワーク・サービス検出方法、ユーザ情報共有方法が含まれる。これらのアルゴリズムに従う装置は、DHCPサービス及びDNSサービスがネットワーク上にまだ存在しない時にはそれらサービスを自動的に開始し、それらサービスが既に存在するならば、その自動開始を行わない。DHCPサービスとDNSサービスは両方が一緒に働いて、人の介入なしに(自動的に)同じネットワーク上のネットワーク装置にネーム及びアドレスを提供する。さらに、サービス情報並びにユーザ情報及びグループ情報を安全に共有する、HTTP(hypertext transfer protocol)ベースの方法が規定される。

【0019】FMAは、NOA装置であるので、DHCP及びDNS、並びにユーザ及びグループ・リストを含む環境設定サービスを必要な時にネットワークに提供することができる。FMAは、他のNOA装置とのHTTPベースの共通インターフェースも提供できる。FMAはネットワークに接続されると、NOAバックアップ装置を検出し、系統的にFMAの存在を他のNOA装置に告知する。

【0020】図1は、FMAを含むネットワーク化システムの一例のブロック図である。図1を参照すると、クライアント110は、ネットワーク100に接続された、電子メール・アプリケーションなどの様々なアプリケーション・プログラムを実行する機能を持つ汎用デジタルコンピュータである。ネットワーク100は、ローカル・エリア・ネットワーク(LAN)であったり、イントラネットであったり、インターネットであったり、あるいは、それ以外の複数装置間で通信可能な任意の相互接続データ経路である。ネットワーク100には、マルチメディア・ソース112、ファクシミリ装置

120、複写機125、プリンタ130、スキャナ135、データ記憶装置140、サーバー145、FMA150も接続されている。一実施例では、FMA150は、図1に示された装置など、ネットワーク100に接続された1つ以上の装置間で伝送されるデータを取り込む機能を備える。

【0021】マルチメディア・ソース112は、ネットワーク100に接続された1つ以上の音声・映像入力装置を意味する。マルチメディア・ソース112は、例えば、アナログデータとデジタルデータの両方をネットワーク100に供給できる1つ以上のマイクロホン及び/又はビデオカメラである。

【0022】ファクシミリ装置120は、ネットワーク100に接続され、電話回線を通じたテキスト及び画像の送受信(ファクス通信)が可能な装置である。ファクシミリ装置120は、一実施例では、印刷形式で作成されたテキストと画像を送信し、別の実施例では、ネットワーク100に接続された任意数の装置によって生成された電子データを送信する。同様に、ファクシミリ装置120は、一実施例では受信データのハードコピーをプリントし、別の実施例では受信データをネットワーク100に接続された任意数の装置へ転送する。

【0023】複写機125は、テキスト及び画像を複製することができる装置である。複写機125は、一実施例では、印刷されたテキストと画像を再生する写真複写機であるが、別の実施例では、ネットワーク100に接続された任意数の装置から受信したデータを再生する写真複写機である。

【0024】プリンタ130は電子データを印刷されたテキスト及び画像へ変換することができる装置であり、スキャナ135は印刷されたテキスト及び画像を電子データへ変換することができる装置である。一実施例では、ファクシミリ装置120、写真複写機125、プリンタ130、スキャナ135はそれぞれ、ネットワーク100に接続された別々の装置である。別の実施例では、これら装置の任意の組み合わせが1台の多機能装置で置き換えられる。なお、本発明の精神及び範囲から逸脱することなく、任意数の装置をネットワーク100から省き、あるいはネットワーク100に追加することができる。

【0025】図1は、ネットワーク100に接続されたデータ記憶装置140も示している。一実施例では、データ記憶装置140は、例えばCD-ROM、DVD-ROM、DVD-RAM、DVD-RW、磁気テープなどの着脱可能な記憶媒体である。別の実施例では、データ記憶装置140はハードディスク・ドライブもしくはは固定ディスク・ドライブのような着脱不可能な記憶媒体である。一実施例では、データ記憶装置140はNOAアーカイブ装置である。

【0026】サーバー145は、ネットワーク100に

(5)

特開2001-155012

9

接続された汎用デジタルコンピュータであり、ネットワーク100に接続された他の装置に対しネットワーク・サービスを提供するように構成される。一実施例では、サーバー145はネットワーク100に対しファイル共有サービスとプリンタ・サービスを提供する。別の実施例では、サーバー145はウェブ・サーバーであり、要求されたHTML (hypertext markup language) ページ又はファイルを、要求元の装置へネットワーク100を通じて提供する。さらに別の実施例では、サーバー145はネットワーク100に対し環境設定サービスを提供できるNOAサーバーである。

【0027】FMA150は、ネットワーク100に接続された、文書取り込みサービスと文書索引付けサービスを提供するファイル管理装置である。一実施例では、FMA150は文書取り込みサービス及び索引付けサービスとともに環境設定サービスをネットワーク100に提供できるNOA装置である。一実施例では、FMA150は、どの装置とも直接接続されず、ネットワーク100を通じて他の装置と通信可能に接続される。FMA150は、NOA装置であるので、その存在をHTTP

プロトコルによってネットワーク100上の他のNOA装置に知らせる機能を有する。  
【0028】文書の取り込み(後により詳しく説明する)は、ある装置がデータ記憶装置140などのアーカイブ装置に文書をアーカイブすることを要求するためのプロセスである。一実施例では、このプロセスは標準のインターネット・プロトコルにより行われる。一実施例では、要求元の装置とアーカイブ装置が共にNOA装置である。他の実施例では、FMA150が要求元装置で

ある。  
【0029】FMA環境においては、文書は単一のファイルからなるであろう。一実施例では、文書という用語が、文書を表すための用語「文書ディレクトリ」と同じ意味で用いられる。文書は、少なくともその内容の一部をなす、色々なタイプの多くの個別ファイルから構成することができる。クライアント110上でプリンタ130に対して生成されるプリント・ジョブは、例えば、サムネイル画像、postscriptファイル、PDF (portable document format) ファイル、及び、抽出されたテキストを含むASCIIファイルとして取り込むことができる。さらに、FMA150は、様々な画像ファイル形式、例えばJPEG形式 (joint photographic experts group format)、GIF形式 (graphics interchange format)、TIFF形式 (tagged image file format) など

10

機能を提供する。例えば、TIFF形式の画像を含むサムネイル画像ファイルは、「ファイル名.thumb.tif」と表現できる。一実施例では、FMA150は、文書のページ番号をファイル名として用いる。このような方法では、文書と同じディレクトリに置かれた複数のファイルで表すことができ、各々のファイルはその文書の、ファイル名で示された別々のページに相当する。例えば、「01.thumb.jpg」はJPEG形式のページ1のサムネイル画像を意味することになる。同様に、「12.thumb.tif」はTIFF形式のページ12のサムネイル画像を意味することになる。

10

【0030】図1を再び参照すると、FMA150は、プリンタ130、ファクシミリ装置120、クライアント110、スキャナ135などのネットワーク100に接続された様々な装置から取り込まれたデータの索引付けを行うことができる。一実施例では、ファクシミリ装置120は電話回線を通じてデータを取り込み、その受信データの少なくとも一部をネットワーク100を介しFMA150へ送信する。別の実施例では、クライアント110からネットワーク100を介しファクシミリ装置120へ送信されたデータは、逐次的に(例えば装置側が知らないうちに)取り込まれ、そのデータの少なくとも一部が索引付けのためにFMA150へ送られる。他の実施例では、ファクシミリ装置120はクライアント110の内部に置かれるため、クライアント110がネットワーク100経由でデータを送信する必要がなくなる。このような実施例でも、FMA150は取り込みデータの少なくとも一部分を受信する。一実施例では、FMA150は文書から抽出された言語的データすなわち「メタデータ」を受信する。一実施例では、ファクシミリ装置120より受信されたデータはTIFF形式に変換されるが、クライアント110から受信されたデータは伝送時の元の形式のままとされる。

10

【0031】このFMAの取り込みプロセスは、スキャナ135や複写機125など、ネットワーク100に接続された他の装置に同様に適用される。一実施例では、スキャン文書又は複写文書に対し光学文字認識(OCR)が実行されるときに、FMA150は2つの特殊なOCR関連ファイルを作成する。一実施例では、FMA150によって、「contents.txt」と「contents.pdf」が生成され、それぞれ文書の全文と返送ページ画像を文書ファイルとして索引付けするために用いられる。

10

【0032】一実施例では、FMA150はネットワーク100上の任意の1つ以上の装置と同じ機能を提供して、それら余分な専用装置を不要にする。しかし、別の実施例では、FMA150は、ネットワーク100経由の文書取り込み及び索引付けをサポートできるだけのハードウェアとソフトウェアを内蔵するシン・サーバーとして実現される。

10

【0033】図2は、FMAのハードウェア構成の一例



(7)

特開2001-155012

11

を示すブロック図である。図2を参照すると、一実施例では、FMA150は制御装置252、入出力装置254、メインメモリ256、記憶装置258、ネットワーク制御装置262を有し、これらは全てバス265を介して接続される。制御装置252は、FMA150の全体的動作を制御するもので、当該技術分野で知られている任意の方法で実現される。一実施例では、制御装置252はデータ処理の分野で知られている汎用マイクロプロセッサからなる。メインメモリ256は、ランダムアクセスメモリ(RAM)、リードオンリーメモリ(ROM)、プログラマブルROM(PROM)、消去可能なPROM(EPROM)、又は、その他の当該技術分野で知られている任意のメモリ・デバイスである。FMA150は、メインメモリ256のほか、二次メモリ装置(不図示)を持つこともできる。一実施例では、FMA150は、それが必要としないUNIXコマンドを削除したUNIXベース・オペレーティングシステムのペアド・ダウン(pared-down)バージョンを利用する。このような実施例では、制御装置252はメインメモリ256より命令を取り出し、解釈し、実行する。一実施例では、FMA150はメインメモリ256内の所定位置にある命令を実行する。

【0034】入出力装置254はバス265に接続され、FMA150への入力とFMA150からの出力を可能にする。入出力装置254は、当該技術分野で知られている任意の汎用入出力装置によって実現される。入出力装置254を、液晶ディスプレイ264やタッチスクリーン274、その他の図2には示さないが当該技術分野で知られている任意の表示装置など、任意台数の表示装置とも接続できる。入出力装置254を、当該技術分野で知られている任意台数のデータ入力装置と接続することもできる。図2は、入出力装置254に接続されたキーボード284とキーパッド294を示しているが、それらに加えて、あるいは、それらに代えて、他の任意の汎用入力装置を接続することができる。一実施例では、図2に示すように、入出力装置254はFMA150の内部にある。別の実施例では、入出力装置254をFMA150の外部に設けられる。

【0035】FMA150のネットワーク200経由の通信を可能にするため、ネットワーク制御装置262がバス265に接続されている。ネットワーク制御装置262は、FMA150のネットワーク接続を可能にするように、ネットワーク・インターフェース・カード、特定用途向け集積回路(ASIC)、その他の当該技術分野で知られている任意の方法で実現できる。一実施例では、図2に示すように、ネットワーク制御装置262はFMA150の内部に設けられる。別の実施例では、ネットワーク制御装置262はFMA150の外部に設けられる。

【0036】記憶装置258はバス265に接続されて

12

おり、当該技術分野で知られている任意の不揮発性記憶装置である。記憶装置258は、CDやDVDなどの着脱可能な記憶媒体を使用してもよいし、ハード/固定ディスクなどの着脱不可能な記憶媒体を使用してもよい。待ち行列ファイル260(後に詳述する)は、記憶装置258上に順序付けられて格納された文書を表す。

【0037】《待ち行列》一実施例では、FMAの動作環境は、1つ以上の処理待ち行列を処理するためのいくつかのプロセス又はデーモンの同時実行を可能にする。

FMAは、文書とも呼ばれるファイルとディレクトリの組み合わせを用いて、そのような待ち行列を作成する。一実施例では、待ち行列ファイル260などの待ち行列ファイルは、改行文字で区切られた文書格納場所のリストを内容としている。

【0038】図3及び図4は、FMA待ち行列処理の一例を示す流れ図である。図3及び図4を参照すると、FMA150は最初に受信待ち行列ファイルのロックの取得を試みる(処理ブロック310)。一実施例では、FMA150は、受信待ち行列ファイルのロックを取得できるか否かを指示するためにメタデータ・ファイル(以下に詳述する)を用いる。FMA150は、受信待ち行列ファイルのロック又は対応したメタデータ・ファイルを取得できないと、所定時間だけ待ち(処理ブロック312)、しかる後に受信待ち行列ファイルのロックの取得を再度試みる(処理ブロック310)。FMA150は、受信待ち行列ファイルのロックを取得できたならば、文書を受信待ち行列ファイルの文書リストに追加する(処理ブロック314)。文書の追加を終えると、FMA150は受信待ち行列ファイルのロックを解除し(処理ブロック316)、受信待ち行列ファイルに処理待ちのエントリがある旨の信号を受信デーモンに送る(処理ブロック318)。

【0039】受信デーモンがそのような受信待ち行列ファイルに文書が存在することを示す信号を検知するか、受信デーモンがそのような信号を検知する前にタイムアウトになると(処理ブロック320)、受信デーモンは受信待ち行列ファイルをロックし(処理ブロック322)、受信待ち行列ファイルに処理すべき文書が実際に存在するか否か調べる(処理ブロック324)。受信待ち行列ファイルに文書が存在しないときには、受信デーモンはスリープ・モードに戻り、再度、受信待ち行列ファイル内に文書があることを示す信号又はタイムアウトの発生を待つ。一方、受信待ち行列ファイルに文書がある場合(処理ブロック324)、受信デーモンは、その文書に“処理中”のマークが付けられているか調べる(処理ブロック326)。その文書に“処理中”マークが付けられているときには、先に実行されたデーモン・プロセスが早く終了した可能性があるため、受信デーモンはその文書を処理せず、その文書に“終了”マークを付ける(処理ブロック328)。一方、その文書に“処

(8)

特開2001-155012

13

理中”マークがついていない場合(処理ブロック326)、受信デーモンはその文書に”処理中”マークを付け(処理ブロック330)、受信待ち行列ファイルのロックを解除し(処理ブロック332)、その文書の通常の処理を続ける(処理ブロック334)。その文書の処理後、受信デーモンはその文書に”終了”マークを付け(処理ブロック336)、受信待ち行列ファイルをロックする(処理ブロック338)。(処理ブロック328、処理ブロック336のいずれであるかを問わず)文書に”終了”マークが付けられると、受信デーモンは、その文書を次の待ち行列に格納し(処理ブロック340)、受信待ち行列から、その文書のエントリを削除する(処理ブロック342)。最後に、受信デーモンは、次の文書処理を可能にするため受信待ち行列ファイルのロックを解除する(処理ブロック344)。

【0040】《メタデータ》一実施例では、FMAシステム内に取り込まれた各文書は、FMA上に、メタデータ・ファイルを含む別々のUNIXディレクトリとして格納される。一実施例では、メタデータ・ファイルには文書に関する特殊な情報、例えば、取り込み装置より得られた言語的データが入っている。一実施例では、文書メタデータは属性とその値のペアからなる。

【0041】図5は、FMAメタデータ・ファイルの一例を示す表である。図5には、文書メタデータの属性が、各々の意味とともに列挙されている。

【0042】図6は、FMAメタデータ・ファイルのもう一つの例を示す表である。図6には、文書メタデータの属性が、そのとり得る値の型とともに列挙されている。

【0043】図7は、XML(extensible markup language)で記述されたFMAメタデータ・ファイルの一例を示す。図7に示した部分的なメタデータ・コードは、ある文書に関し生成されるもの、すなわち、その文書がユーザ”jones”により取り込まれたこと(行640)、ユーザ”jones”の所有であること(行630)、あるNOAによりバックアップされたこと(行690)を説明している。さらに、図7に示したメタデータ・コードは、その文書がOCR及び圧縮で処理されたこと(行660及び行670)、索引付けされたこと(行680)、ディスク37にバックアップされたこと(行690)を表している。

【0044】FMA又はNOAが、(http://WWW.W3.orgのワールド・ワイド・ウェブ・コンソーシアム(W3C)より入手可能なXML仕様書で規定された)ウェルフォームド(well-formed)ファイルでないメタデータ・ファイルと遭遇したときには、一実施例では、そのFMA又はNOAは、そのメタデータを、その文書について検証可能な全ての情報を含むウェルフォームド・ファイルで置き換える。その結果、ウェル

14

フォームドであるが属性/値のペアを含まない全く空のメタデータ・ファイルとなる可能性がある。一実施例では、図5、図6及び図7に示したメタデータ・ファイル及びデータ・フィールド定義はXML仕様に従って作られるが、同様の機能を得るために、他のプログラム言語、例えばHTML、SGML(standard generalized markup language)、その他の当該技術分野で知られている任意の言語を同様に使用することができる。一実施例では、メタデータ・ファイルに対し項目を追加又は削除するため、FMAは、小さなソフトウェア・ライブラリもしくはルーチンを使って、メタデータ・ファイルを解析するとともに、文書又はメタデータの項目に関するロック操作、読み出し操作、書き込み操作及び削除操作を行う。

【0045】FMAは、FMA上の文書を管理するマスター文書リストも保持するであろう。一実施例では、マスター文書リストは、FMAでこれまでに取り込まれた全ての文書に関する情報を指示する唯一の情報源である。一実施例では、マスター文書リストは、図5、図6及び図7に示されたメタデータ・ファイルに見られるものと同じ属性を保持する。マスター文書リストを作るために、内容索引付けが可能な任意のデータベースを利用することができる。

【0046】《文書の取り込み》文書の取り込みは、NOA装置がアーカイブNOA装置に文書のアーカイブを要求する動作である。FMA環境内で、NOA装置に取り込みを要求する任意のネットワーク・エンティティ又は装置によって文書取り込みを実行することができる。一実施例では、文書取り込みはFMAシステム内の全文書に対し自動的に実行されるが、別の実施例では、ユーザによる設定に応じて、選択された文書のみが取り込まれる。例えば、ユーザは、文書の自動取り込みを防止するボタンを押すことができる。一実施例では、文書取り込みに標準的なインターネット・プロトコルを用い、複数文書の取り込みあるいは1つの文書に関連した複数ファイルの取り込みを可能にする。

【0047】一実施例では、HTTPが取り込みプロトコルとして用いられる。HTTPのPOST操作により、FMAで、例えばアーカイブ形式を含む様々な形式のファイルを取り込むことができる。このようなアーカイブ・ファイルは別々の名前のデータ・ファイルの集合を収容でき、また、テープ・アーカイブ(tar)形式、zip形式、gzip形式、pkzip形式、Stuffit形式などのアーカイブ形式をとることができる。

【0048】別の実施例では、FTP(Internet File Transfer Protocol)のインプリメンテーションが取り込みプロトコルとして用いられることにより、文書はデフォルトのFTPディレクトリにMIME(multipurpose Internet mail extension)ファイルとして、あるいは

15

は同デフォルト・ディレクトリのサブディレクトリとして取り込まれる。文書の名前はNOAのディレクトリ・システムの一部としては格納されず、メタデータ・ファイルに格納されるため、文書ディレクトリの実際の名前は文書取り込みでは重要でない。一実施例では、文書の取り込みデータが文書ディレクトリの名前として用いられる。また、連続番号のようなユニークな識別子を各文書に割り当て、文書のメタデータ・ファイルに格納する（図5、図6及び図7に、420、520、620として各々示されている）。

【0049】図8は、FMA環境の一実施例において用いられるMIME取り込みの一例を示す流れ図である。図8を参照すると、NOA取り込み装置は、1つ以上のコンテンツ・ファイルと1つのメタデータ・ファイルを含むMIMEマルチパート・ファイルを生成する（処理ブロック710）。NOA取り込み装置は、そのMIMEファイルを生成した後、デスティネーションのNOA装置との匿名FTPセッションの確立を試みる（処理ブロック715）。FTPセッションの確立後、NOA取り込み装置はデスティネーションNOA装置上でユニークなファイル名を決定し（処理ブロック720）、そのファイルのデスティネーションNOA装置への転送を試みる（処理ブロック725）。この転送を失敗した場合、取り込み装置は別のファイル名を取得して、ファイル転送を再度試みる。ファイル転送が成功すると（処理ブロック730）、取り込みは完了する。

【0050】図9は、FMA環境の一実施例におけるディレクトリ取り込みの一例の流れ図である。図9を参照すると、取り込み装置はデスティネーションNOAとの匿名FTPセッションを確立する（処理ブロック735）。このFTPセッションが確立した後、取り込み装置はデスティネーションNOA上でユニークであると指定されるディレクトリ名（以下に詳述する）を決定する（処理ブロック740）。ディレクトリ名の決定後、取り込み装置は、デスティネーションNOA上に、その名前を持つディレクトリの作成を試みる（処理ブロック745）。ディレクトリ名の重複、その他の理由で、このディレクトリ作成の試みが成功しないときには、取り込み装置は別のディレクトリ名を決定し、ディレクトリの作成を再度試みる（処理ブロック750、740、745）。一方、取り込み装置は、デスティネーションNOA上にディレクトリを首尾良く作成できると（処理ブロック750）、1つ又は複数のコンテンツ・ファイルを、新たに作成したディレクトリに転写する（処理ブロック750）。取り込み装置はメタデータ・ファイルも生成し（処理ブロック760）、このメタデータ・ファイルがFMA装置へ送られ（処理ブロック765）、プロセスが完了する。

【0051】図10は、FMA環境での文書格納プロセスの一実施例の流れ図を示す。文書格納プロセス中に、

(9)

特開2001-155012

16

FMA又はNOAは適当なディレクトリを作成し、そのディレクトリに文書を移動させ、マスター・リストを更新する。図10を参照すると、格納しようとする文書のメタデータ・ファイルがアクセスされ、その“Capture date”フィールド（図5及び図6の項目450、550を各々参照）が取り出される（処理ブロック770）。その文書の“Capture date”フィールドが存在しないか、そのメタデータ・ファイルさえも存在しないときには、現在のシステム時刻が取得され、それが文書の“Capture

date”として用いられる（処理ブロック774）。一方、その文書の“Capture date”が存在するときには、システムは適切な名前が付けられたディレクトリが存在するか調べる。一実施例では、文書ディレクトリは“yy/mm/dd”で表される。ここで、yyは文書の作成年、mmは文書の作成月、ddは文書の作成日を意味する。

【0052】図10を再度参照する。システムは、適切な4桁年で表されたディレクトリが存在するか調べる（処理ブロック776）。適切な年を表すディレクトリが存在しないときには、システムはそのようなディレクトリを作成する（処理ブロック778）。一方、適切な年を表すディレクトリが存在するときには、システムは、その年ディレクトリ内に適切な月を表すディレクトリが存在するか調べる（処理ブロック780）。その年ディレクトリ内に適切な月ディレクトリが存在しないときには、システムは、その年ディレクトリ内に月ディレクトリを作成する（処理ブロック782）。適切な年ディレクトリと月ディレクトリが存在するときには、システムは最後に、そのネストされた年/月ディレクトリ内に適切な日ディレクトリが存在するか調べる（処理ブロック784）。その日ディレクトリが存在しないときには、システムはその年/月ディレクトリ内に適切な日ディレクトリを作成する（処理ブロック786）。一方、適切な年月日を表すディレクトリが既に存在するならば、システムは文書が格納される新しい文書ディレクトリ名を生成する。一実施例では、システムは、既存の文書ディレクトリ名の最後に付けられる4桁の乱数を生成する（処理ブロック788）。ユニークな文書ディレクトリ名が確定したならば（処理ブロック786、788）、文書がそのディレクトリに移され（処理ブロック790）、その文書の新たな格納場所を反映するようにマスター文書リストが更新される（処理ブロック792）。

【0053】《ウェブ・インターフェースとユーザ・インターフェースの具体例》FMAは、多様なグラフィカル入力・表示ページを使ってFMAユーザ・インターフェースを表現する。また、FMAは、ネットワーク接続を經由したデータ送受信を可能にするウェブ・ベースのグラフィカル・インターフェースを備える。

【0054】FMAのウェブ・インターフェースを介して、ユーザはFMAに取り込まれた文書を検索して取り

50

(10)

特開2001-155012

17

出すことができる。一実施例では、FMAは図5及び図6に示したようなメタデータ属性を用いて、時間優先問い合わせと内容優先問い合わせの両方による取り込み文書の一元的な検索を可能にする。一実施例では、最初の問い合わせによって見つかった任意の文書集合に関し、引き続き問い合わせをすることができる。したがって、ユーザは、内容優先問い合わせの結果に基づいて時間優先問い合わせを発することができ、また、これと逆に、時間優先問い合わせの結果に基づいて内容優先問い合わせを発することができる。また、ユーザは、新しい問い合わせを発するよりも、前の問い合わせを修正したいと思うこともあるかもしれない。一実施例では、ユーザは、前に提示されたフィールド・エントリを含む前の問い合わせフォームに戻って、適当な検索条件を入力して問い合わせを修正することができる。

【0055】FMAは、問い合わせをFMA内に常駐する独立文書として保存し、時間優先又は内容優先の検索条件を使って以前保存した問い合わせを検索する機能を有する。さらに、FMAは、ユーザが、1ページ又は複数ページに表示された問い合わせ結果を含め、選択した文書を見て見るために、ある格納場所すなわち「スタック」に追加できるようにするための機構を提供する。一実施例では、FMAは、文書をスタックに追加するため、文書メタデータ・ファイルに含まれている各文書の識別子又は連続番号をスタック内容ファイルの最後に書き込む。一実施例では、保存された問い合わせがスタックに追加されるときには、その問い合わせを表すHTMLページがスタックに追加される。文書問い合わせ結果の表示方法と同様の方法で、例えば後述のリスト表示で、スタックの内容を表示させることができる。

【0056】一実施例では、FMAは、特定の問い合わせ結果を表示するための1つ以上のウェブ・ページを、CGIスクリプトを用いて生成する。FMA問い合わせ結果は、例えばリスト表示、カレンダー表示などの様々な時間優先表示、あるいは、例えばストリーム表示やサムネイル表示などの様々な内容優先表示によって、グラフィカルに表示できる。一実施例では、ユーザは、文書表示設定に応じて、各々の問い合わせ結果表示を選択して切り替えることができる。

【0057】本発明のリスト表示によれば、多数の文書に関する大量のテキスト情報をユーザに伝えることができる。図12は、リスト表示のFMA問い合わせ結果ページの一例を示す。図12では、「relative sentences」という問い合わせの条件に合致した文書のリストが表示されている。このリスト表示においては、例えば、参照番号902で示されている、生成ソース（文書を生じた、又は取り込んだ装置）、生成時刻、ファイル・サイズ、検索結果スコアなどの様々な属性によって問い合わせ結果をソートすることができる。ソース・アイコン904は、該当文書が生成された又は取り込まれた装

18

置をユーザにグラフィカルに表示する。一実施例では、個々の文書の取り込みソースは、その文書のメタデータ・ファイル内に特定されている（図5及び図6では、参照番号445、545でそれぞれ表されている）。一実施例では、ユーザがソース・アイコン又は対応した文書をダブルクリックすると、その文書がFTP経由でユーザのマシンに取り出される。一実施例では、カレンダー表示は、問い合わせ結果を標準的な月間カレンダーに似た形に整形して表示する。カレンダー表示は、個々の問い合わせ結果のイベント・ベースの表示をユーザに提供し、文書程度の移動指標も兼ねる。一実施例では、取り込み文書を表すアイコンは、その文書が取り込まれた日に対応したカレンダー上の位置にグラフィカル表示される。このような実施例では、ある日について表示されるアイコンの数は、その日に取り込まれた文書の数に相当する。別の実施例では、文書又は文書を表すアイコンを、文書の取り込み日ではなく生成日に対応したカレンダー上の位置に表示させ、あるいは、生成日と取り込み日の両方に対応した位置に表示させることができる。

【0058】図13は、カレンダー表示のFMA問い合わせ結果ページの一例を示す。一実施例では、ユーザがカレンダー表示中の文書ソース・アイコンを指示すると、その文書のサムネイル・バージョンが画面上に挿入画像として表示される。同様に、ユーザが、あるアイコンを例えばダブルクリックして選択すると、その文書がダウンロードされる。

【0059】本発明のストリーム表示は、取り込み時刻順に文書を並べたグラフィカルな文書ストリームの形で問い合わせ表示する。一実施例では、問い合わせ結果として得られた各文書の第1ページとその取り込み日が、各文書を互いに少しだけずらして並べた形でグラフィカル表示される。一実施例では、取り込み日はハイパーリンクであり、それが選択された時に、その日に取り込まれた文書を含む新たなページをFMAに表示させる。一実施例では、補助的な文書表示も用意され、ストリーム表示上で文書が選択されると、その文書の拡大バージョンがその要約情報とともに表示画面上の付随ウインドウ内に表示される。図14は、ストリーム表示のFMA問い合わせ結果ページの一例を示す。一実施例では、サムネイル表示は上記ストリーム表示と似がよっている。一実施例では、サムネイル表示は、問い合わせ結果として得られた各文書の第1ページをサムネイル画像としてグラフィカルに表示する。一実施例では、サムネイル画像は、サムネイル表示中に並べて表示される。図15は、サムネイル表示のFMA問い合わせ結果ページの一例を示す。

【0060】《文書の編集》FMA上の文書の所有者は、その文書に関し、表示、所有権の変更、削除、他の所有者への複写など様々な操作を行うことができる。所有者が特定の文書に対して持つ支配権は広範であるた

(11)

特開2001-155012

19

め、FMA上のファイルに対するアクセスは“単独所有者”の所有権モデルで管理される。すなわち、すべての文書は、どのユーザ又はユーザたちにアクセス又は閲覧が許可されるか指定する所有者フィールドを含んでいる。FMA上の所有者は、個人ユーザとして指定するかユーザ・グループとして指定することができる。あるグループが文書の所有者として指定されたときには、そのグループの各メンバーがそのファイルの所有者として扱われる。一実施例では、FMA上の各文書は公開文書に指定することができ、そうすると、文書の所有権は変更されないが全てのユーザがその文書の閲覧を許される。一実施例では、所有権アクセス用のユーザ・グループ名スペースは、ウェブ・サーバーに対しHTTP認証を与えるために使用されるユーザ・グループ名スペースと同じである。

【0061】《文書フローの具体例》図11は、FMA150のデーモンと待ち行列の間の文書フローの一例を示す。図11を参照すると、入力デーモン845、FTPデーモン842、プリント取り込みデーモン844、OCR/抽出デーモン855、索引付けデーモン865、圧縮デーモン875、バックアップ・デーモン885、HTTPデーモン895は全て一連の待ち行列により結合される。待ち行列は、デーモンに処理すべき文書が存在することを通知する。入力デーモン845は、ファイル管理システムによって取り込まれた各文書の初期処理を管理するデータ取り込みマネージャである。最初に、入力デーモン845は2つの待ち行列ディレクトリ、すなわち入力待ち行列340と処理待ち行列848を監視し、その2つのディレクトリの内容に対する処理タスクを呼び出す。ファイル管理システムの取り込みインターフェースを提供するFTPデーモン842により、又は、プリントされる各プリント・ジョブのコピーを逐次的に取り込むプリント取り込みデーモン844によって、入力待ち行列840に文書ディレクトリが満たされる。入力デーモン845は、これらの文書を検出し、それを処理待ち行列848へ転送する。処理待ち行列848に文書がある間、入力デーモン845は定期的にそれらの文書をチェックして別のタスクを読み出し可能であるか判定する。

【0062】入力デーモン845は、タスクを呼び出す時に、処理待ち行列848から圧縮待ち行列870、索引付け待ち行列860、OCR/抽出待ち行列850、バックアップ待ち行列880のいずれかに文書を送る。圧縮待ち行列870は文書を圧縮デーモン875に送り、圧縮デーモン875は記憶スペースの節約のためにその文書のコンテンツ・ファイルの圧縮を試み、そして圧縮した文書を処理待ち行列848へ送る。一実施例では、圧縮デーモン875は、未加工のビットマップはMMR圧縮により圧縮するが、それ以外のコンテンツ・ファイルはzipフォーマットにより圧縮する。索引付け

20

待ち行列860は文書を索引付けデーモン865に送り、索引付けデーモン865はその文書の索引付けを行い、索引付けした文書を処理待ち行列848へ送り返す。OCR/抽出待ち行列850は文書をOCR/抽出デーモン855へ送り、OCR/抽出デーモン855はその文書のテキストに対する文字認識を行い、またその文書から画像データを抽出する。その後、OCR/抽出デーモン855は、処理した文書を処理待ち行列848へ送り返す。

【0063】バックアップ待ち行列880は、ファイル管理システムでバックアップする必要のある文書のリストを示すファイルである。バックアップ待ち行列880は多くのプロセスに共有される。バックアップ・デーモン885は、バックアップ待ち行列880と結合され、文書のバックアップ、バックアップ保留、削除及び索引ファイル管理に関する多くの機能を遂行する。一実施例では、バックアップ・デーモン885は、いくつかの独立したプロセスとして構築される。あるプロセスが実際のバックアップ動作を管理し、もう1つのプロセスが非同期的な文書復旧を管理し、3つ目のプロセスが非同期的な文書削除を管理する。

【0064】バックアップ・デーモン885とHTTPデーモン895の間に、復旧待ち行列892と削除待ち行列890がある。復旧待ち行列892は装置によりハードディスク上に復旧させる必要のある文書のリストであり、削除待ち行列890は装置により削除する必要のある文書のリストである。HTTPデーモン895は、削除待ち行列890及び復旧待ち行列892上の文書及び/又は文書ステータスを表示するためのウェブ・インターフェース（例えばページ）を提供する。HTTPデーモン895は、ファイル管理システムの他の待ち行列上の文書及び/又は文書ステータスの表示のためのウェブ・インターフェースも提供することができる。

【0065】《信頼性の強化》データ故障、さらにはデータ喪失の危険を減らすべく、FMAはバックアップ・エンジン及び再起動デーモンなどのFMAベース・システムの信頼性を向上させる様々な機構を実装している。

【0066】FMAバックアップ・エンジンは、文書のバックアップ、バックアップ保留、削除、及び索引ファイルに関連した多くの機能を遂行するバックアップ・デーモンを含む。一実施例では、FMA文書は可能な限り速やかにバックアップされる。すなわち、FMA文書バックアップ要求は非同同期に受け付けられることによって、FMAは最も“重要な”要求を真先に処理する。要求の重要度はFMAによってヒューリスティックに決定される。同様に、FMA文書の全体及び部分を、1つ以上のバックアップ媒体から非同同期に復旧し得る。ここに述べたバックアップ・復旧プロセスの具体例に関するこれ以上の情報は“Method and Apparatus for Opportunistic Queue Processing”の表題で1998年11月1

2日に出願され、本願出願人に譲渡され、かつ、ここに援用される同時係属出願第09/191,277号に見出すことができる。

【0067】FMAシステムは、前述の非同期のファイル・バックアップ手順及び復旧手順に加えて、並列オフサイト・バックアップ・ディスクの書き込み機能も装備している。FMAは、オフサイト・バックアップ中に、必要に応じて外部へ持ち出し可能に設計された二次バックアップ・ディスクを生成する。一実施例では、このFMAオフサイト・バックアップは、ユーザの直接操作の結果として動作する一時的なプログラムである。一実施例では、このFMAオフサイト・バックアップ・ルーチンは、リムーバブル記憶媒体に、全ての文書をバックアップし終わるまで、あるいは記憶媒体が一杯になるまで、ファイルを順次複写する。

【0068】FMAシステムは、ハードディスク並びに各種バックアップ装置の内容によってマスター文書リスト・データベースを復旧する機能も備わっている。一実施例では、各FMA文書は、マスター文書リスト・データベースの復旧を可能にするためバルク・サブミット (bulk submit) ファイルを使用する。各々のバルク・サブミット・ファイルは、特定の文書に関する文書トランザクションの完全なリストを、それに対応した各トランザクション発生時のタイムスタンプとともに格納している。FMAは、これらのバルク・サブミット・ファイルを保持し、マスター文書リスト・データベースのレコードが変更される度に、それらバルク・サブミット・ファイルがバックアップされたことを確認する。マスター文書リスト・データベースの復旧が行われる度に、対応したバルク・サブミット・ファイルのタイムスタンプが参照されて、そのバルク・サブミット・ファイル又はマスター文書リスト・データベースにもっと新しいデータが入っていないか確認される。

【0069】FMAベースのシステムの信頼性向上のために設計された、その種のもう1つの機構が再起動デーモンである。再起動デーモンは、他のデーモン・プログラム又はプロセスの機能を検査し、それらのプロセスを、それが障害を起こすか応答しなくなった場合に再起動するマスター・プロセスである。一実施例では、再起動デーモン自体の機能を検査し、その障害時に再起動デーモンを再度初期化する特別なプロセスが実装される。

【0070】なお、ここで遂行される処理は処理ロジックによって実行されるが、この処理ロジックはハードウェア (例えば、専用ロジック、回路など)、ソフトウェア、あるいは、その両者の組み合わせから構成することができる。

【0071】以上の説明を読めば、本発明の多くの変形及び修正が当業者に明白となるように、説明のために図示及び記述したいずれの実施例も決して限定を意図したものではないことを理解すべきである。したがって、様々

な実施例に関し詳細に述べた内容は、特許請求の範囲を限定することを意図するものではない。

【0072】

【発明の効果】以上の説明から理解されるように、本発明によれば、様々なソースによって生成された文書を管理するためのユーザの介入を減らし、また、効率的な文書検索のための単純明解なインターフェースを提供できるなどの効果を得られるものである。

【図面の簡単な説明】

10 【図1】FMAを含むネットワーク化システムの構成例を示す。

【図2】FMAのハードウェア構成例を示す。

【図3】FMAの待ち行列処理フローの一例を示す。

【図4】図3に示した待ち行列処理フローの続きを示す。

【図5】FMAメタデータ・ファイル形式の一例を示す。

【図6】FMAメタデータ・ファイル形式の別の一例を示す。

20 【図7】XML形式のFMAメタデータ・ファイルの一例を示す。

【図8】FMA環境におけるMIME取り込みの処理フローの一例を示す。

【図9】FMA環境におけるディレクトリ取り込みの処理フローの一例を示す。

【図10】FMA環境における文書格納プロセスの処理フローの一例を示す。

【図11】FMAのデーモンと待ち行列の間の文書フローの一例を示す。

30 【図12】リスト表示のFMA問い合わせ結果ページの一例を示す。

【図13】カレンダー表示のFMA問い合わせ結果ページの一例を示す。

【図14】ストリーム表示のFMA問い合わせ結果ページの一例を示す。

【図15】サムネイル表示のFMA問い合わせ結果ページの一例を示す。

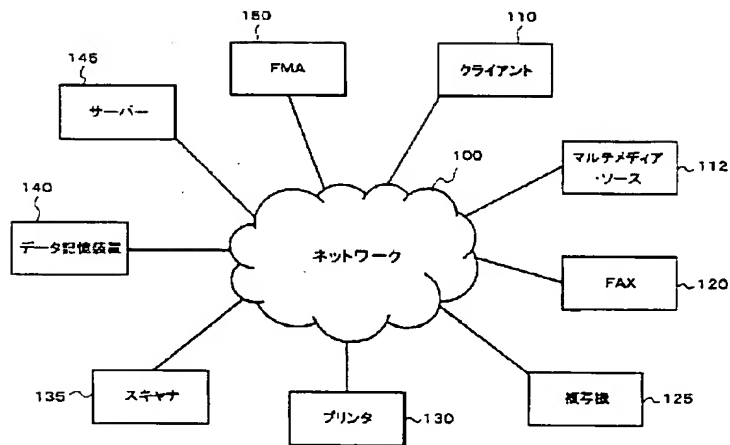
【符号の説明】

100	ネットワーク
110	クライアント
112	マルチメディア・ソース
120	ファクシミリ装置
125	複写機
130	プリンタ
135	スキャナ
140	データ記憶装置
145	サーバー
150	FMA (ファイル管理装置)
200	ネットワーク
50 252	制御装置

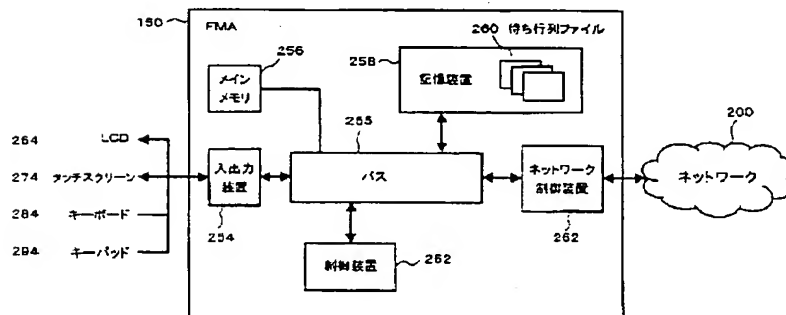
23  
 254 入出力装置  
 256 メインメモリ  
 258 記憶装置  
 262 ネットワーク制御装置  
 265 バス  
 260 待ち行列ファイル  
 840 入力待ち行列  
 842 FTPデーモン  
 844 プリント取り込みデーモン  
 845 入力デーモン  
 848 処理待ち行列

24  
 \* 850 OCR／抽出待ち行列  
 855 OCR／抽出デーモン  
 860 索引付け待ち行列  
 865 索引付けデーモン  
 870 圧縮待ち行列  
 875 圧縮デーモン  
 885 バックアップ・デーモン  
 890 削除待ち行列  
 892 復旧待ち行列  
 10 895 HTTPデーモン  
 \*

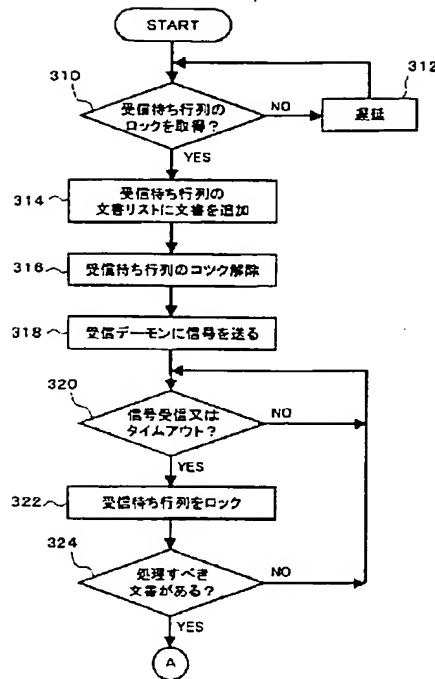
【図1】



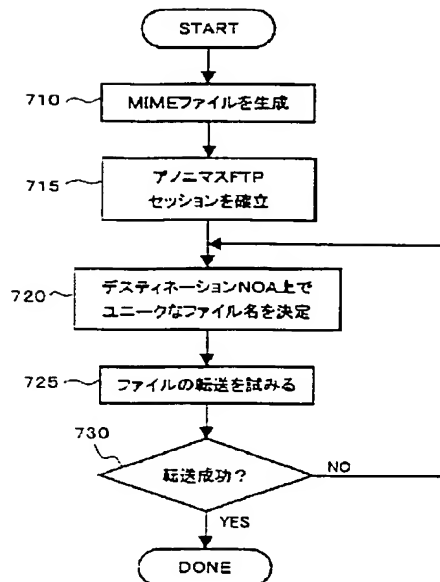
【図2】



【図3】



【図8】

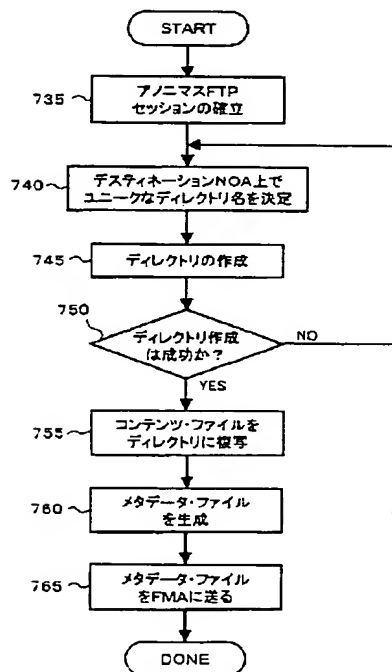


【図7】

```

<NOA_metadata
<doc
610   <item name="FMA ID" value = "2361738294" /
620   <item name="serial number" value = "883639" /
630   <item name="document owner" value = "jones" /
640   <item name="document capturer" value = "jones" /
650   <item name="capture date" value = "03/29/98 01:17:45 GMT" /
660   <item name="OCRed" value = "true" /
670   <item name="compressed" value = "true" /
680   <item name="indexed" value = "true" /
690   <item name="backed up" value = "37" /
</doc
</NOA_metadata
  
```

【図9】

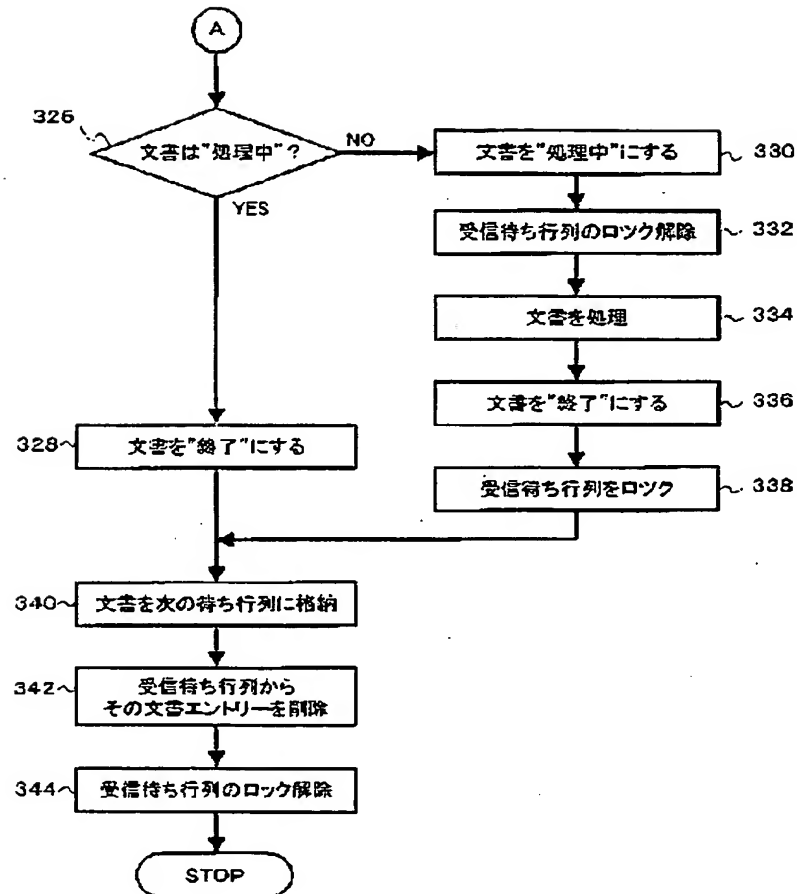




(15)

特開2001-155012

【図4】



(15)

特開2001-155012

【図5】

	属性名	説明
410	FMA ID	FMAのための世界中でユニークな識別子
420	Serial number	NOA上でユニークな文書の識別子
	Document name	任意に定められた文書の名前
430	Document owner	著作権を現在所有しているユーザ/グループ
440	Document capture	文書を取り込んだユーザ/グループ
445	Capture device type	文書が取り込まれたデバイスの種類
	Capture device name	文書が取り込まれたデバイスの名称
	Transport direction	文書が、その取り込み時において、発信文書であったのか、受信文書であったのか、そのどちらでもなかったのか、その両方であったのか
450	Capture date	文書がアーカイブされた日付
	Creation date	文書が取り込みNOAによって取り込まれた日付又はPC上で最後に修正された日付
	Last read on	写しが発行に読まれた日付
	Location	メイン・ディスク/バックアップ・ディスク上の格納位置
460	Backed Up	文書がバックアップされたか否かと、文書がバックアップされたバックアップ媒体を示す識別子
	Public	文書が公開文書か否か
480	Indexed	文書の索引付けの有無
	Total pages	文書の総ページ数(判明している場合)
470	Compressed	圧縮デモンにより文書が処理済みか否か
460	OCRed	文書が光学文字認識済みか否か
	Return receipt to	索引付けの後に受領通知が返送されるURL
	Fax number	文書がファクス文書の場合の相手側ファクス番号
	Mail from	文書が電子メールの場合の?from?ヘッダ・アドレス
	Mail to	文書が電子メールの場合の?to?ヘッダ・アドレス
	File to index	索引付け対象のファイルの名前
	Thumbnail	サムネイル・ファイルの名前
	Copy of	文書が索引付け後に転写された場合には、そのオリジナル文書のserial number、そうでない場合には、その文書自体のserial number

(17)

特開2001-155012

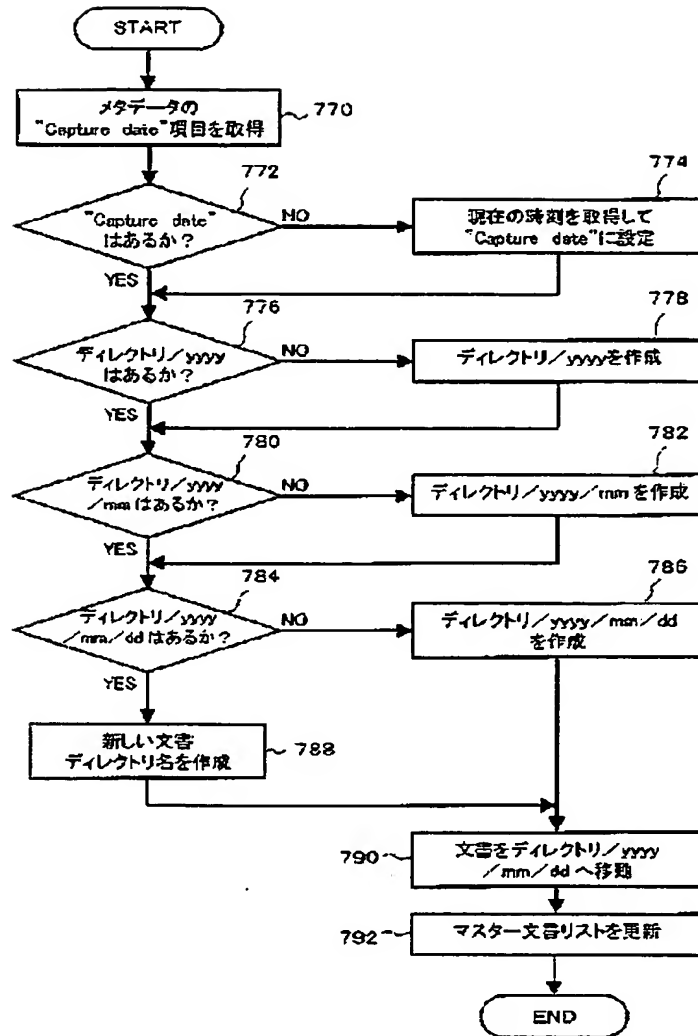
【図6】

	属性名	値の型
510	FMA ID	正整数を表す10進数文字列
520	Serial number	正整数を表す10進数文字列
	Document name	任意の文字データ
530	Document owner	登録されたユーザ/グループの名前を表す文字列、又は、複数のユーザ/グループをコンマで区切って表したリスト
540	Document capture	登録されたユーザ/グループの名前を表す文字列
545	Capture device type	取り込みNOAの種類を表す文字列、"copier"、"Fax"、"Printer"、"email"、"PC"を選択できる
	Capture device name	取り込みNOAのネットワーク名を表す文字列
	Transport direction	"in"、"out"、"no"、"both"のいずれか
550	Capture date	UNIX日付を表す文字列で、一般にGMTのタイムゾーンでは"YYYYMMDDhhmmss"の形式
	Creation date	UNIX日付を表す文字列で、一般にGMTのタイムゾーンでは"YYYYMMDDhhmmss"の形式
	Last read on	UNIX日付を表す文字列で、一般にGMTのタイムゾーンでは"YYYYMMDDhhmmss"の形式
	Location	ディレクトリのパス名を表す文字列
590	Backed Up	非負整数を表す10進数文字列、0は文書がバックアップされていないことを意味し、それ以外の数はバックアップ後の識別子を意味する
	Public	ブール文字列で、"true"、"false"のいずれか
580	Indexed	ブール文字列で、"true"、"false"のいずれか
	Total pages	非負整数を表す10進数文字列、0は文書の総ページ数が分からないことを意味し、それ以外の数は総ページ数を意味する
570	Compressed	ブール文字列
560	OCRed	ブール文字列
	Return receipt to	URLを表す文字列
	Fax number	10進数文字列
	Mail from	e-mailアドレスを表す文字列
	Mail to	e-mailアドレスを表す文字列
	File to index	ディレクトリのないファイル名を表す文字列
	Thumbnail	File to index と同じ
	Copy of	Serial number と同じ

(18)

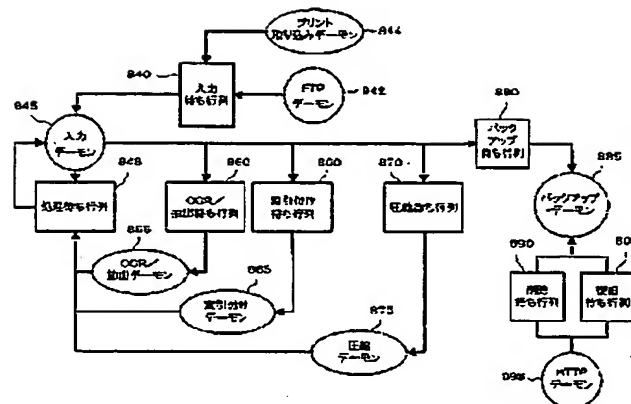
特開2001-155012

【図10】

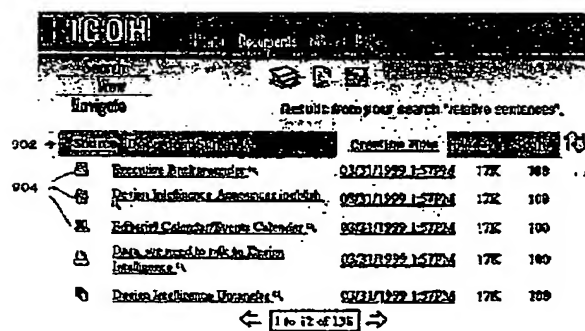


特開2001-155012

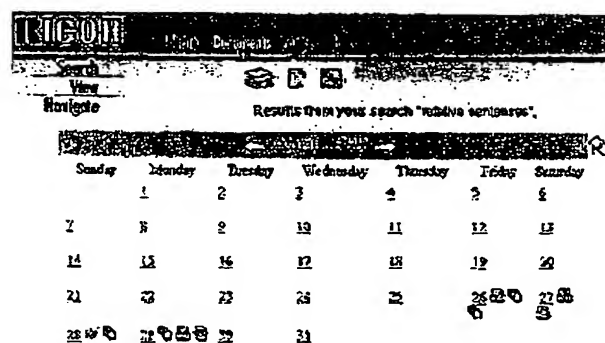
४७५



【圖 12】



【图 13】

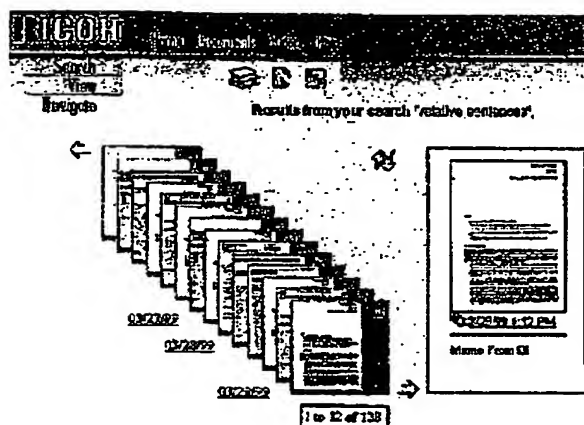


**BEST AVAILABLE COPY**

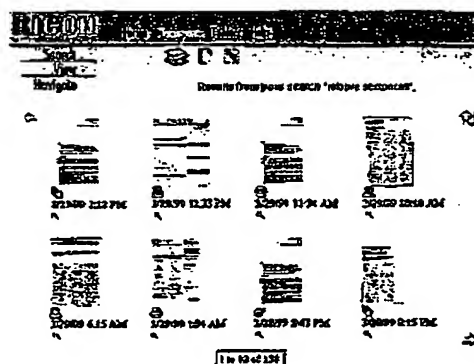
(20)

特開2001-155012

【図14】



【図15】



BEST AVAILABLE COPY